An extension of the Ky Fan inequality

Yuri Suhov*, **Salimeh Yasaei Sekeh**†

Abstract

The aim of this paper is to analyze the weighted KyFan inequality proposed in [11]. A number of numerical simulations involving the exponential weighted function is given. We show that in several cases and types of examples one can imply an improvement of the standard KyFan inequality.

Key words: weight function, weighted KyFan inequality, maximizing, weighted conditional and mutual entropies, weighted exponential function

2000 MSC: 60A10, 60B05, 60C05

1 Introduction. The weighted Ky Fan inequality

1.1. The well-known Ky Fan inequality [6, 7, 8] asserts that $\log\det\mathbf{C}$ is a concave function of a (strictly) positive definite matrix $\mathbf{C}$. In other words, $\forall$ strictly positive-definite $d \times d$ matrices $\mathbf{C}_1$, $\mathbf{C}_2$ and $\lambda_1, \lambda_2 \geq 0$ with $\lambda_1 + \lambda_2 = 1$,

$$\log\det\Big(\lambda_1\mathbf{C}_1 + \lambda_2\mathbf{C}_2\Big) - \sum_{a=1,2}\lambda_a \log\det\mathbf{C}_a \geq 0;\ \text{equality iff } \lambda_1\lambda_2 = 0 \text{ or } \mathbf{C}_1 = \mathbf{C}_2. \tag{1.1}$$

For original 'geometric' proofs of (1.1) and other related inequalities, see Ref [10] and the bibliography therein. In [3, 5, 4] the derivation of (1.1) occupies few lines and is based on properties of information-theoretical entropies; a similar method allows to derive a number of other determinant-related inequalities.

*Math Dept, Penn State University, PA 16802, USA; DPMMS, University of Cambridge, CB30WB, UK; IITP, RAS, 127994 Moscow GSP-4, Russia. E-mail: ims14@ps.edu, yms@statslab.cam.ac.uk

†Department of Statistics, Federal University of São Carlos (UFSCar), São Carlos, Brazil. E-mail: sa_yasaei@yahoo.com

More precisely, (1.1) is equivalent to the bound for Shannon differential entropies for Gaussian probability density functions (PDFs):

$$h(f^{\rm No}_{\lambda_1\mathbf{C}_1+\lambda_2\mathbf{C}_2}) - \lambda_1 h(f^{\rm No}_{\mathbf{C}_1}) - \lambda_2 h(f^{\rm No}_{\mathbf{C}_2}) \geq 0;\ \text{equality iff } \lambda_1\lambda_2 = 0 \text{ of } \mathbf{C}_1 = \mathbf{C}_2. \tag{1.2}$$

Here and below, $f^{\rm No}_{\mathbf{C}}$ stands for the d-variate normal PDF $\mathrm{N}(\mathbf{0}, \mathbf{C})$, with mean $\mathbf{0}$ and covariance matrix $\mathbf{C}$:

$$f^{\rm No}_{\mathbf{C}}(\mathbf{x}_1^d) := \frac{1}{(2\pi)^{d/2}\big(\det\mathbf{C}\big)^{1/2}} \exp\left(-\frac{1}{2}{\mathbf{x}_1^d}^{\rm T}\mathbf{C}^{-1}\mathbf{x}_1^d\right),\ \ \mathbf{x}_1^d = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^d. \tag{1.3}$$

Next, $h(f) = -\displaystyle\int_{\mathbb{R}^d} f(\mathbf{x}_1^d)\log\, f(\mathbf{x}_1^d){\rm d}\mathbf{x}_1^d$ represents the Shannon differential entropy of a PDF f. In the Gaussian case,

$$h(f^{\rm No}_{\mathbf{C}}) := -\int_{\mathbb{R}^d} f^{\rm No}_{\mathbf{C}}(\mathbf{x}_1^d \log\, f^{\rm No}\mathbf{C}(\mathbf{x}_1^d){\rm d}\mathbf{x}_1^d = \frac{1}{2}\log\Big[(2\pi)^d(\det\mathbf{C})\Big] + \frac{d\log\, e}{2}. \tag{1.4}$$

Inequality (1.2) is a consequence of the fact that, under certain conditions, $h(f)$ is maximized $f = f^{\rm No}_{\mathbf{C}}$. Throughout the paper, we use the abbreviation KFI for either of (1.1), (1.2). (Sometimes the term a standard KFI is also used.)

1.2. In this paper we compare inequalities (1.1), (1.2) with *weighted* inequalities similar to (1.2) and established for *weighted* differential entropies in the recent paper [11]; see below. For the sake of pre-emptiveness, we call each of these inequalities a *weighted* Ky Fan inequality (WKFI, for short). A WKFI is obtained for a given non-negative weight function; when this function equals 1, the WKFI coincides with the KFI. A natural question is whether a WKFI can provide an 'improvement' to KFI; for instance, by producing a positive lower bound for the LHS in (1.1), (1.2). We give a numerical evidence that the answer can be yes or no, depending on the choice of $\mathbf{C}_a$ and λ_a. We work with so-called exponential weight functions for which all calculations simplify. Furthermore, the numerical simulations are done for $d = 2$, allowing a graphical representation of results.

Let $\mathbf{x}_1^d \in \mathbb{R}^d \mapsto \phi(\mathbf{x}_1^d) \geq 0$ be a given non-negative measurable function positive on an open domain in $\mathbb{R}^d$. Following [1], [2] and [11], under the usual agreement, $0\cdot\log\, 0 = 0\cdot\log\,(+\infty) = 0$, the *weighted* differential entropy (WDE) of PDF f with *weight function* (WF) ϕ is defined by

$$h^{\rm w}_\phi(f) := -\int_{\mathbb{R}^d} \phi(\mathbf{x}_1^d)f(\mathbf{x}_1^d)\log\, f(\mathbf{x}_1^d){\rm d}\mathbf{x}_1^d, \tag{1.5}$$

assuming that the integral is absolutely convergent. Cf. [1, 2, 9].

In the Gaussian case, the WDE $h^{\rm w}_\phi(f^{\rm No}_{\mathbf C})$ admits a representation extending the RHS in (1.4). Define a number $\alpha(\mathbf C)=\alpha_\phi(\mathbf C)>0$ and a $d\times d$ matrix $\mathbf\Phi_{\mathbf C}=\mathbf\Phi^{\rm No}_{\phi,\mathbf C}$ involving WF ϕ and PDF $f^{\rm No}_{\mathbf C}$:

$$\alpha(\mathbf C)=\int\limits_{\mathbb R^d}\phi(\mathbf x_1^d)f^{\rm No}_{\mathbf C}(\mathbf x_1^d){\rm d}\mathbf x_1^d,\;\;\mathbf\Phi^{\rm No}_{\mathbf C}=\int\limits_{\mathbb R^d}\mathbf x_1^d\left(\mathbf x_1^d\right)^{\rm T}\phi(\mathbf x_1^d)f^{\rm No}_{\mathbf C}(\mathbf x_1^d){\rm d}\mathbf x_1^d. \tag{1.6}$$

Then

$$h^{\rm w}_\phi(f^{\rm No}_{\mathbf C})=\frac{\alpha(\mathbf C)}{2}\log\Big[(2\pi)^d(\det\mathbf C)\Big]+\frac{\log\,e}{2}{\rm tr}\,\mathbf C^{-1}\mathbf\Phi_{\mathbf C}:=\sigma_\phi(\mathbf C). \tag{1.7}$$

1.3. The following theorem was proven in [11].

Theorem 1.1 (The WKFI; cf. [11], Theorem 3.2). *Given* $\lambda_1,\lambda_2\in[0,1]$ *with* $\lambda_1+\lambda_2=1$ *and (strictly) positive-definite* $\mathbf C_1$*,* $\mathbf C_2$*, set:* $\mathbf C=\lambda_1\mathbf C_1+\lambda_2\mathbf C_2$ *and* $\mathbf\Psi=\lambda_1\mathbf\Phi_{\mathbf C_1}+\lambda_2\mathbf\Phi_{\mathbf C_2}-\mathbf\Phi_{\mathbf C}$. *Assume that, ofr a given WF* ϕ*,*

$$\begin{array}{c}\lambda_1\alpha(\mathbf C_1)+\lambda_2\alpha(\mathbf C_2)-\alpha(\mathbf C)\geq 0,\\ \Big[\lambda_1\alpha(\mathbf C_1)+\lambda_2\alpha(\mathbf C_2)-\alpha(\mathbf C)\Big]\log\Big[(2\pi)^d(\det\mathbf C)\Big]+{\rm tr}\,\Big(\mathbf C^{-1}\mathbf\Psi\Big)\leq 0.\end{array} \tag{1.8}$$

Then

$$\sigma_\phi(\lambda_1\mathbf C_1+\lambda_2\mathbf C_2)-\lambda_1\sigma_\phi(\mathbf C_1)-\lambda_2\sigma_\phi(\mathbf C_2)\geq 0;\;\;\textit{equality iff}\;\;\lambda_1\lambda_2=0\;\;\textit{or}\;\;\mathbf C_1=\mathbf C_2. \tag{1.9}$$

Observe that when $\phi(x)\equiv 1$, bounds (1.8) are fulfilled for all choices of $\mathbf C_a$ and λ_a, $a=1,2$. (In fact, they become equalities.) In this case, inequality (1.9) coincides with (1.2).

2 Exponential weight functions

2.1. As was said, in this paper we deal with *exponential* WFs, of the form

$$\phi(\mathbf x_1^d)=\exp\left({\mathbf t_1^d}^{\rm T}\mathbf x_1^d\right)\;\;\text{where }\mathbf t_1^d\in\mathbb R^d. \tag{2.1}$$

To shorten the notation, we write from now on $\mathbf x$ and $\mathbf t$ instead of $\mathbf x_1^d$ and $\mathbf t_1^d$. Here we use the Laplace transform formulas: for $\phi(\mathbf x)=\exp\left(\mathbf t^{\rm T}\mathbf x\right)$ the Eqn (1.6) yields

$$\alpha(\mathbf C)=\exp\left(\frac12\mathbf t^{\rm T}\mathbf C\mathbf t\right)\;\;\text{and}\;\;\Phi_{\mathbf C}=\mathbf C\exp\left(\frac12\mathbf t^{\rm T}\mathbf C\mathbf t\right). \tag{2.2}$$

Hence, for $\phi(\mathbf x)=\exp\left(\mathbf t^{\rm T}\mathbf x\right)$, the WDE (1.7) becomes

$$\sigma_{\mathbf t}(\mathbf C)=h(f^{\rm No}_{\mathbf C})\exp\left(\frac12\mathbf t^{\rm T}\mathbf C\mathbf t\right) \tag{2.3}$$

where $h(f^{\rm No}_{\mathbf{C}})$ is as in Eqn (1.4).

To tackle condition (1.8), we introduce the set $\mathbb{S} = \mathbb{S}(\mathbf{C}_1, \mathbf{C}_2; \lambda_1, \lambda_2) \subset \mathbb{R}^d$:

$$\begin{aligned}
&\mathbb{S} = \Big\{ \mathbf{t} \in \mathbb{R}^d :\ F^{(1)}(\mathbf{t}) \geq 0,\ \textit{and}\ F^{(2)}(\mathbf{t}) \leq 0 \Big\}\ \textit{where} \\
&\quad F^{(1)}(\mathbf{t}) = \sum_{\alpha=1,2} \lambda_\alpha \exp\left(\frac{1}{2}\mathbf{t}^{\rm T}\mathbf{C}_\alpha\mathbf{t}\right) - \exp\left(\frac{1}{2}\mathbf{t}^{\rm T}\mathbf{C}\mathbf{t}\right)\ \textit{and} \\
&\quad F^{(2)}(\mathbf{t}) = \left[\sum_{\alpha=1,2} \lambda_\alpha \exp\left(\frac{1}{2}\mathbf{t}^{\rm T}\mathbf{C}_\alpha\mathbf{t}\right) - \exp\left(\frac{1}{2}\mathbf{t}^{\rm T}\mathbf{C}\mathbf{t}\right)\right] \log\Big[(2\pi)^d(\det\mathbf{C})\Big] \\
&\qquad\qquad + \sum_{\alpha=1,2} \lambda_\alpha \exp\left(\frac{1}{2}\mathbf{t}^{\rm T}\mathbf{C}_\alpha\mathbf{t}\right) {\rm tr}\left[\mathbf{C}^{-1}\mathbf{C}_\alpha\right] - d\exp\left(\frac{1}{2}\mathbf{t}^{\rm T}\mathbf{C}\mathbf{t}\right).
\end{aligned} \tag{2.4}$$

Theorem 1.1 is transformed into Theorem 2.1:

Theorem 2.1 *Given positive definite matrices* $\mathbf{C}_1$, $\mathbf{C}_2$ *and* $\lambda_1, \lambda_2 \in [0,1]$ *with* $\lambda_1 + \lambda_2 = 1$, *set* $\mathbf{C} = \lambda_1\mathbf{C}_1 + \lambda_2\mathbf{C}_2$. *Assume that* $\mathbf{t} \in \mathbb{S}$. *Then*

$$h(f^{\rm No}_{\mathbf{C}}) \exp\left(\frac{1}{2}\mathbf{t}^{\rm T}\mathbf{C}\mathbf{t}\right) - \lambda_1 h(f^{\rm No}_{\mathbf{C}_1}) \exp\left(\frac{1}{2}\mathbf{t}^{\rm T}\mathbf{C}_1\mathbf{t}\right) - \lambda_2 h(f^{\rm No}_{\mathbf{C}_1}) \exp\left(\frac{1}{2}\mathbf{t}^{\rm T}\mathbf{C}_2\mathbf{t}\right) \geq 0; \tag{2.5}$$

equality iff $\lambda_1\lambda_2 = 0$ *or* $\mathbf{C}_1 = \mathbf{C}_2$.

2.2. Consequently, $\forall\ \mathbf{t} \in \mathbb{S}$ we have that

$$\begin{aligned}
\Sigma(\mathbf{t}) := &\Big\{\log\Big[(2\pi e)^d(\det\mathbf{C})\Big]\Big\} \exp\left(\frac{1}{2}\mathbf{t}^{\rm T}\mathbf{C}\mathbf{t}\right) \\
&- \sum_{a=1,2} \lambda_a \Big\{\log\Big[(2\pi e)^d(\det\mathbf{C}_a)\Big]\Big\} \exp\left(\frac{1}{2}\mathbf{t}^{\rm T}\mathbf{C}_a\mathbf{t}\right) \geq 0.
\end{aligned} \tag{2.6}$$

Eqn (2.6) can be called an exponentially weighted (or briefly: an exponential) Ky Fan inequality (EKFI). When $\mathbf{t} = 0$, the EKFI is reduced to the standard KFI (1.1), (1.2). In particular, for $\mathbf{t} = 0$, the bounds in (2.4) become equalities for any $\mathbf{C}_a$ and λ_a. (Hence, point $\mathbf{t} = \mathbf{0}$ lies in $\mathbb{S}$ for any choice of $\mathbf{C}_a$ and λ_a.) Consequently, it makes sense to analyze the difference $\Lambda(\mathbf{t}) := \Sigma(\mathbf{t}) - \Sigma(\mathbf{0})$:

$$\begin{aligned}
\Lambda(\mathbf{t})(= \Lambda(\mathbf{t}; \mathbf{C}_1, \mathbf{C}_2; \lambda_1, \lambda_2)) = &\log(\det\mathbf{C})\left[\exp\left(\frac{1}{2}\mathbf{t}^{\rm T}\mathbf{C}\mathbf{t}\right) - 1\right] \\
&+ \sum_{\alpha=1,2} \lambda_\alpha \log(\det\mathbf{C}_\alpha)\left[1 - \exp\left(\frac{1}{2}\mathbf{t}^{\rm T}\mathbf{C}_\alpha\mathbf{t}\right)\right] \\
&+ d\log(2\pi e)\left[-\sum_{\alpha=1,2} \lambda_\alpha \exp\left(\frac{1}{2}\mathbf{t}^{\rm T}\mathbf{C}_\alpha\mathbf{t}\right) + \exp\left(\frac{1}{2}\mathbf{t}^{\rm T}\mathbf{C}\mathbf{t}\right)\right].
\end{aligned} \tag{2.7}$$

If $\Lambda(\mathbf{t}) > 0$, we can think of an improvement in the standard KFI, and if $\Lambda(\mathbf{t}) < 0$, of a deterioration.

2.3. It has to be said that set $\mathbb{S}$ looks rather involved, and there is no guaranty that it is not empty. (In fact, numerical evidence suggests that $\mathbb{S} = \emptyset$ for some choices of $\mathbf{C}_a$ and λ_a.) Therefore, it makes sense to explore the behavior of $\Lambda(\mathbf{t})$ for $\mathbf{t}$ in the whole of $\mathbb{R}^d$. In particular, a stationary point $\mathbf{t}$ satisfying $\mathrm{grad}_{\mathbf{t}}\Lambda = 0$ is found from

$$\begin{aligned} 0 = d\log(2\pi e)&\left[\sum_{\alpha=1,2}\lambda_\alpha\mathbf{C}_\alpha\mathbf{t}\exp\left(\frac{1}{2}\mathbf{t}^{\mathrm{T}}\mathbf{C}_\alpha\mathbf{t}\right) - \mathbf{C}\mathbf{t}\exp\left(\frac{1}{2}\mathbf{t}^{\mathrm{T}}\mathbf{C}\mathbf{t}\right)\right] \\ + \sum_{\alpha=1,2}\lambda_\alpha\log\left(\det\mathbf{C}_\alpha\right)&\mathbf{C}_\alpha\mathbf{t}\left[\exp\left(\frac{1}{2}\mathbf{t}^{\mathrm{T}}\mathbf{C}_\alpha\mathbf{t}\right)\right] - \log\left(\det\mathbf{C}\right)\mathbf{C}\mathbf{t}\left[\exp\left(\frac{1}{2}\mathbf{t}^{\mathrm{T}}\mathbf{C}\mathbf{t}\right)\right] \end{aligned} \tag{2.8}$$

or

$$\begin{aligned} \mathbf{t} = \sum_{\alpha=1,2}\lambda_\alpha\mathbf{C}^{-1}\mathbf{C}_\alpha\mathbf{t}&\left\{\exp\left[\frac{1}{2}\mathbf{t}^{\mathrm{T}}(\mathbf{C}_\alpha - \mathbf{C})\mathbf{t}\right]\right\} \\ &\times\frac{d\log e + \log\left[(2\pi)^d(\det\mathbf{C}_\alpha)\right]}{d\log\, e}. \end{aligned} \tag{2.9}$$

An obvious solution is $\mathbf{t} = \mathbf{0}$; we are tempting to suggest that it is unique. To analyze the character of this point, let us take the second gradient:

$$\begin{aligned} \nabla^2_{\mathbf{tt}}\Lambda(\mathbf{t}) = d\log(2\pi e)&\left[\sum_{\alpha=1,2}\lambda_\alpha(\mathbf{C}_\alpha + \mathbf{C}_\alpha\mathbf{t}\mathbf{t}^{\mathrm{T}}\mathbf{C}_\alpha^{\mathrm{T}})\exp\left(\frac{1}{2}\mathbf{t}^{\mathrm{T}}\mathbf{C}_\alpha\mathbf{t}\right)\right. \\ &\left.-(\mathbf{C} + \mathbf{C}\mathbf{t}\mathbf{t}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}})\exp\left(\frac{1}{2}\mathbf{t}^{\mathrm{T}}\mathbf{C}\mathbf{t}\right)\right] \\ + \sum_{\alpha=1,2}\lambda_\alpha\log&\left(\det\mathbf{C}_\alpha\right)(\mathbf{C}_\alpha + \mathbf{C}_\alpha\mathbf{t}\mathbf{t}^{\mathrm{T}}\mathbf{C}_\alpha^{\mathrm{T}})\exp\left(\frac{1}{2}\mathbf{t}^{\mathrm{T}}\mathbf{C}_\alpha\mathbf{t}\right) \\ &- \log\left(\det\mathbf{C}\right)(\mathbf{C} + \mathbf{C}\mathbf{t}\mathbf{t}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}})\exp\left(\frac{1}{2}\mathbf{t}^{\mathrm{T}}\mathbf{C}\mathbf{t}\right). \end{aligned} \tag{2.10}$$

At $\mathbf{t} = 0$ it yields

$$\nabla^2_{\mathbf{tt}}\Lambda(\mathbf{t})\Big|_{\mathbf{t}=\mathbf{0}} = \sum_{\alpha=1,2}\lambda_\alpha\log\left(\det\mathbf{C}_\alpha\right)\mathbf{C}_\alpha - \log\left(\det\mathbf{C}\right)\mathbf{C}. \tag{2.11}$$

For a local minimum at the origin $\mathbf{t} = \mathbf{0}$ we need the matrix $\sum\limits_{\alpha=1,2}\lambda_\alpha\log\left(\det\mathbf{C}_\alpha\right)\mathbf{C}_\alpha - \log\left(\det\mathbf{C}\right)\mathbf{C}$ to be positive definite. In other words, the following property emerges, featuring reduced convexity of the map $\mathbf{C}\mapsto\mathbf{C}\log\det\mathbf{C}$: for given positive definite $\mathbf{C}_1$, $\mathbf{C}_2$,

$$\begin{aligned} \lambda\in[0,1]\mapsto[\lambda\mathbf{C}_1 + (1-\lambda)\mathbf{C}_2]\log\,\det\left[\lambda\mathbf{C}_1 + (1-\lambda)\mathbf{C}_2\right] \\ \text{is a convex matrix-valued function.} \end{aligned} \tag{2.12}$$

Again, we can say that our numerical evidence suggests that $\mathbf{t} = \mathbf{0}$ can be a local extremum (of either type) or a saddle point.

2.4. Let us check the status of the origin in the case $d = 1$. Here $\mathbf{t} = t \in \mathbb{R}$, and $\mathbf{C}_a = c_a$, $a = 1, 2$, and $\mathbf{C} = c = \lambda_1 c_1 + \lambda_2 c_2$ are scalars. Further, function $\Lambda(t)$ from Eqn (2.7) has the form

$$\begin{aligned}\Lambda(t) = \log(2\pi e)&\left[\sum_{\alpha=1,2} \lambda_\alpha \exp\left(\frac{1}{2}c_\alpha t^2\right) - \exp\left(\frac{1}{2}ct^2\right)\right]\\ &+ \sum_{\alpha=1,2} \lambda_\alpha (\log\, c_\alpha)\left[\exp\left(\frac{1}{2}c_\alpha t^2\right) - 1\right] - (\log\, c)\left[\exp\left(\frac{1}{2}ct^2\right) - 1\right].\end{aligned} \tag{2.13}$$

Next, functions $F^{(1)}$ and $F^{(2)}$ from Eqn (2.4) become

$$\begin{aligned}F^{(1)}(t) &= \sum_{\alpha=1,2} \lambda_\alpha \exp\left(\frac{1}{2}c_\alpha t^2\right) - \exp\left(\frac{1}{2}ct^2\right),\\ F^{(2)}(t) &= \left[\sum_{\alpha=1,2} \lambda_\alpha \exp\left(\frac{1}{2}c_\alpha t^2\right) - \exp\left(\frac{1}{2}ct^2\right)\right]\log\,(2\pi c)\\ &\quad + \sum_{\alpha=1,2} \lambda_\alpha \exp\left(\frac{1}{2}c_\alpha t^2\right)c^{-1}c_\alpha - \exp\left(\frac{1}{2}ct^2\right).\end{aligned} \tag{2.14}$$

Correspondingly, for set $\mathbb{S}$ we obtain:

$$\begin{aligned}\mathbb{S} = \Bigg\{&t \in \mathbb{R}:\; \lambda_1 \exp\left[\frac{1}{2}\lambda_2(c_1 - c_2)t^2\right] + \lambda_2 \exp\left[\frac{1}{2}\lambda_1(c_2 - c_1)t^2\right] \geq 1,\\ &\left(\lambda_1 \exp\left[\frac{1}{2}\lambda_2(c_1 - c_2)t^2\right] + \lambda_2 \exp\left[\frac{1}{2}\lambda_1(c_2 - c_1)t^2\right] - 1\right)\log\,(2\pi c)\\ &\quad + \frac{\lambda_1 c_1}{c}\exp\left[\frac{1}{2}\lambda_2(c_1 - c_2)t^2\right] + \frac{\lambda_2 c_2}{c}\exp\left[\frac{1}{2}\lambda_1(c_2 - c_1)t^2\right] \leq 1\Bigg\}.\end{aligned} \tag{2.15}$$

It is can be seen that if $0 < c < 1/(2\pi)$, set $\mathbb{S}$ is unbounded.

The stationary point is where $\dfrac{\partial}{\partial t}\Lambda = 0$ or

$$\begin{aligned}0 = t\,\log(2\pi e)&\left[\sum_{\alpha=1,2} \lambda_\alpha c_\alpha \exp\left(\frac{1}{2}c_\alpha t^2\right) - c\exp\left(\frac{1}{2}ct^2\right)\right]\\ &+t \sum_{\alpha=1,2} \lambda_\alpha c_\alpha (\log\, c_\alpha)\exp\left(\frac{1}{2}c_\alpha t^2\right) - c(\log\, c)\exp\left(\frac{1}{2}ct^2\right).\end{aligned}$$

Clearly, $t = 0$ is a solution. Next, we calculate the second derivative $\dfrac{\partial^2}{\partial t^2}\Lambda$ at $t = 0$. A general expression is:

$$\begin{aligned}\frac{\partial^2}{\partial t^2}\Lambda(t) = \log(2\pi e)&\left[\sum_{\alpha=1,2} \lambda_\alpha c_\alpha (1 + c_\alpha t^2)\exp\left(\frac{1}{2}c_\alpha t^2\right) - c(1 + ct^2)\exp\left(\frac{1}{2}ct^2\right)\right]\\ + \sum_{\alpha=1,2} \lambda_\alpha c_\alpha &(1 + c_\alpha t^2)(\log\, c_\alpha)\exp\left(\frac{1}{2}c_\alpha t^2\right) - c(1 + ct^2)(\log\, c)\exp\left(\frac{1}{2}ct^2\right).\end{aligned}$$

Taking into account that $\sum_{\alpha=1,2} \lambda_\alpha c_\alpha = c$, we obtain that

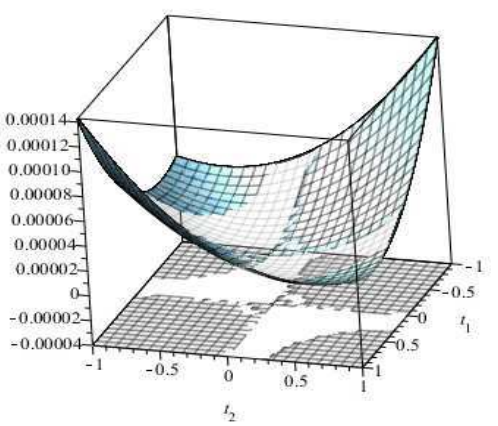
$$\frac{\partial^2}{\partial t^2}\Lambda(t)\Big|_{t=0} = \sum_{\alpha=1,2} \lambda_\alpha c_\alpha (\log\, c_\alpha) - c \log\, c \geq 0. \tag{2.16}$$

Inequality (2.16) holds since $x \mapsto x \log\, x$ is a convex function for $x > 0$.

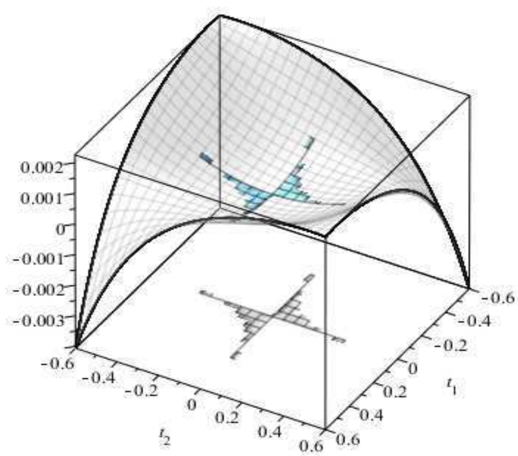
3 Numerical results

Performing numerical simulations, we show the graph of function $\mathbf{t} \mapsto \Lambda(\mathbf{t})$ in Eqn (2.7) for $d = 2$ within chosen ranges of argument $\mathbf{t} = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$ around point $\mathbf{t} = \mathbf{0}$. Matrices $\mathbf{C}_a$ are taken in the form $\begin{pmatrix} \sigma_a^2 & \rho_a \sigma_a^2 \\ \rho_a \sigma_a^2 & \sigma_a^2 \end{pmatrix}$ where $\sigma_a > 0$ and $|\rho_a| < 1$.

The simulations show the behavior of function $\Lambda(\mathbf{t})$ for chosen matrices $\mathbf{C}_1$ and $\mathbf{C}_2$ and values of $\lambda_1 = \lambda$ and $\lambda_2 = 1 - \lambda$ over selected ranges of the argument $\mathbf{t} = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$ and indicate the set $\mathbb{S}$. The plots exhibit a variety of possible patterns: positive and negative values of $\Lambda(\mathbf{t})$, convexity, concavity, global/local minimum/maximum, as well as a saddle point, at $\mathbf{t} = \mathbf{0}$. (Recall, $\Lambda > 0$ has been proposed as an improvement whereas $\Lambda < 0$ as a retrogression of a standard KFI.)



A: $\sigma_1 = 0.1$, $\sigma_2 = 0.1001$, $\rho_2 = -0.666$, $\lambda = 0.99$



B: $\sigma_1 = 1$, $\sigma_2 = 1.06$, $\rho_1 = 0.5$, $\rho_2 = -0.2$, $\lambda = 0.01$

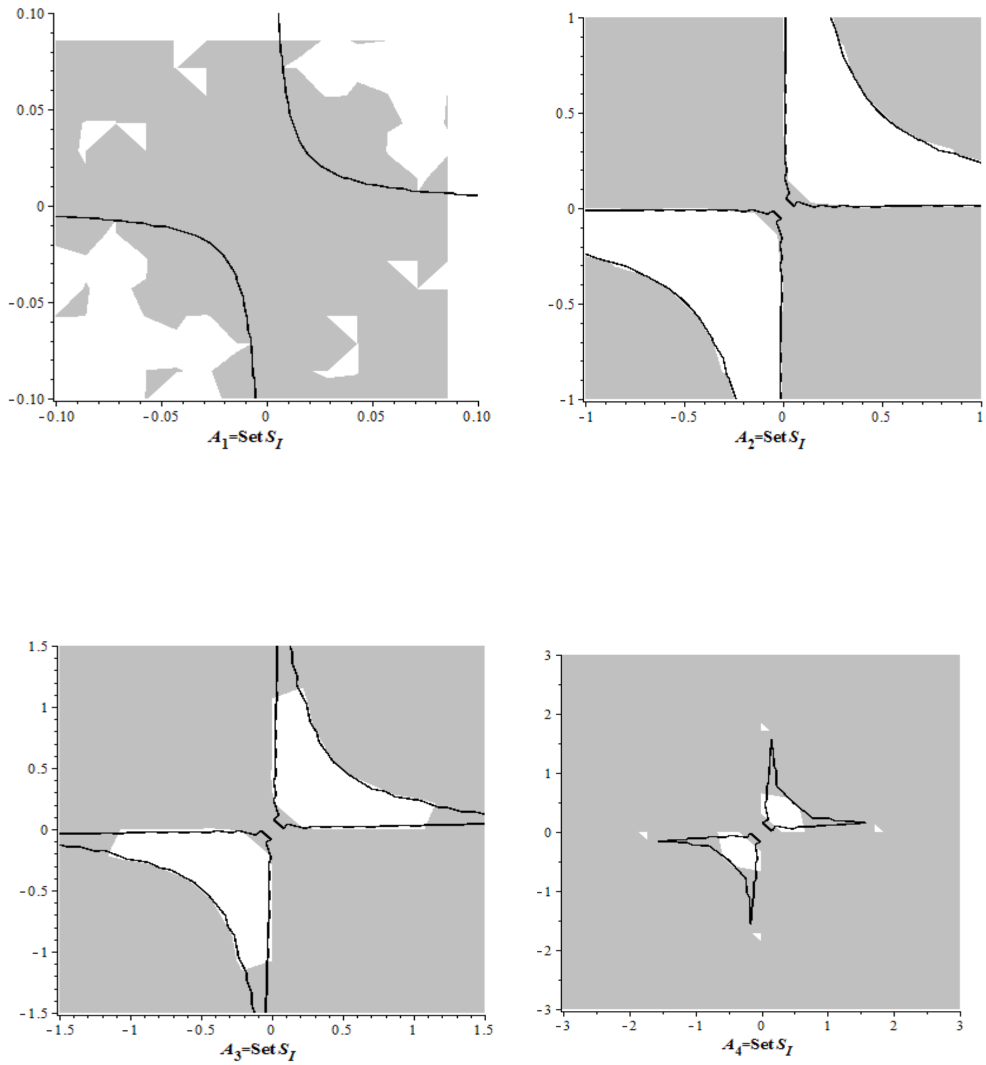
Figure 3.1

In Fgures 3.1 – 3.5, set $\mathbb{S}$ – when it is non-empty – is shown in a grey color at a bottom horizontal plane. (The level at which this plane is placed has been selected for presentational

convenience only.) We want to note that in some examples the origin $\mathbf{t} = \mathbf{0}$ seems to be an isolated point in $\mathbb{S}$: it may be a consequence of the fact that the WF $\phi(\mathbf{x}) = \exp\left(\mathbf{t}^{\mathrm{T}}\mathbf{x}\right)$ is unbounded for $\mathbf{t} \neq \mathbf{0}$.

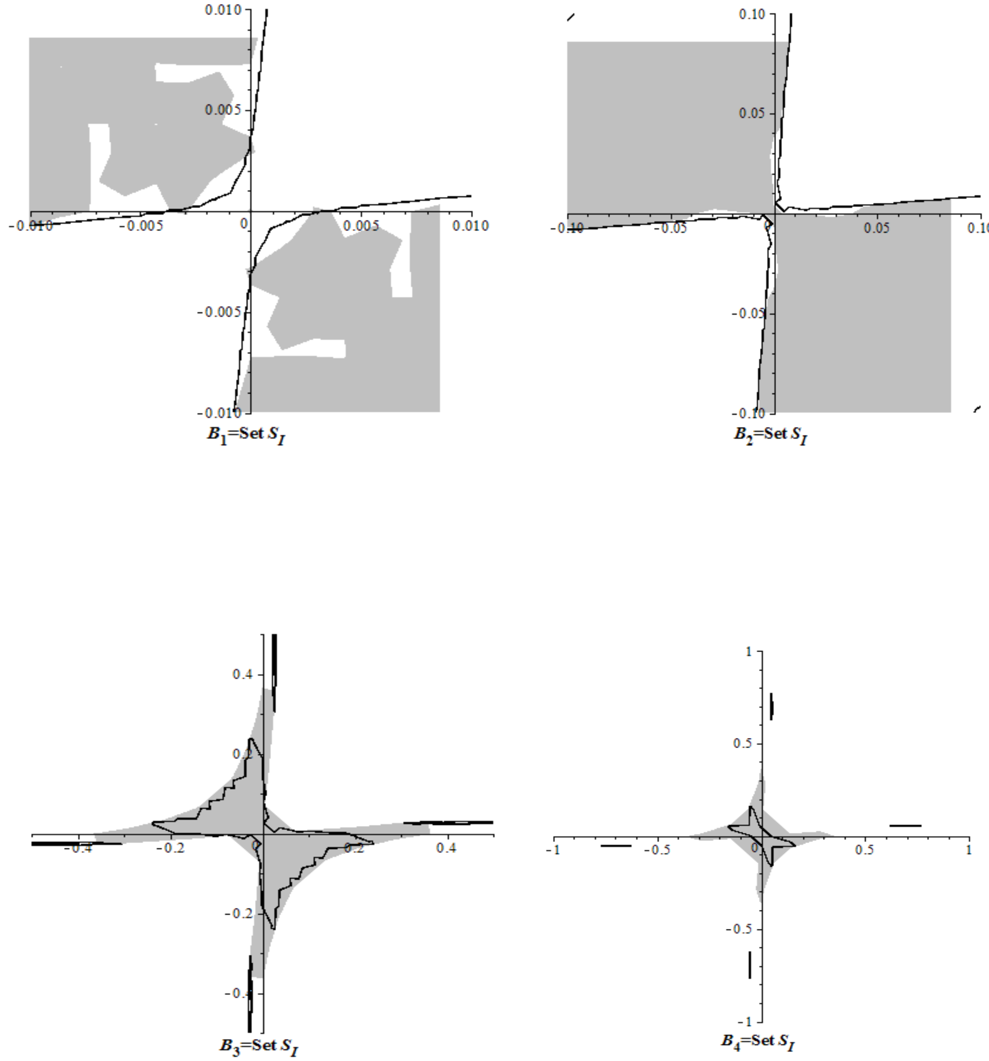
In Figures 3.1.A and 3.1.B the graph of function $\Lambda(\mathbf{t})$ lies above the value 0 attained at $\mathbf{t} = \mathbf{0}$. This suggests that $\mathbf{t} = \mathbf{0}$ is a global minimum of $\Lambda(\mathbf{t})$. Consequently, the matrix in the RHS of Eqn (2.11) is positive definite for the specified choices of $\mathbf{C}_1$, $\mathbf{C}_2$ and λ.

Apparently, the EKFI holds true far beyond $\mathbb{S}$ and yields an improvement of the standard KFI. Matrix $\mathbf{C}_1$ in Figure 3.1.A is of the form $\mathbf{C}_1 = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_1^2 \end{pmatrix}$ and $\mathbf{C}_2 = \begin{pmatrix} \sigma_2^2 & \sigma_2^2\rho_2 \\ \sigma_2^2\rho_2 & \sigma_2^2 \end{pmatrix}$. The values of $\lambda = \lambda_1$ are chosen to be 0.99 on Figure 3.1.A and 0.001 on Figure 3.1.B.



The plots above give an impression of set $\mathbb{S}$ in Figure 3.1.A showing how our perception changes when the range of variables t_1 and t_2 increases. In this example, the set (shown in the

grey color) is, obviously, unbounded. (The hyperbola-type curves are used to provide a geometric reference.) On the other hand, the plots below demonstrate that set $\mathbb{S}$ in Figure 3.1.B is bounded.



Next, Figure 3.2 shows a more complex character of behavior. Here $\mathbf{t} = \mathbf{0}$ is, apparently, a saddle point for the graph of $\Lambda(\mathbf{t})$. Function $\Lambda(\mathbf{t})$ takes both positive and negative values. However, over set $\mathbb{S}$ the EKFI yields an improvement of the standard KFI. Here matrices $\mathbf{C}_\alpha = \begin{pmatrix} \sigma_\alpha^2 & \sigma_\alpha^2\rho_\alpha \\ \sigma_\alpha^2\rho_\alpha & \sigma_\alpha^2 \end{pmatrix}$, $\alpha = 1, 2$, and $\lambda_1 = \lambda_2 = \lambda = 1/2$.

Further, a similar pattern of behavior is confirmed on Figures 3.3.A and 3.3.B, with the same form of matrices $\mathbf{C}_\alpha$, and with $\lambda_1 = \lambda = 0.99$ and 0.5, respectively.

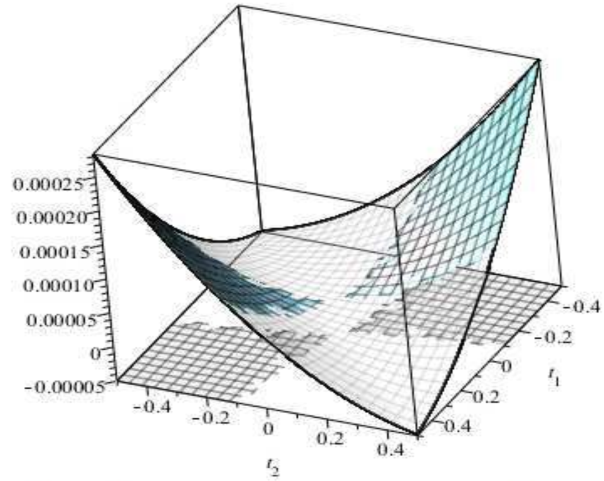
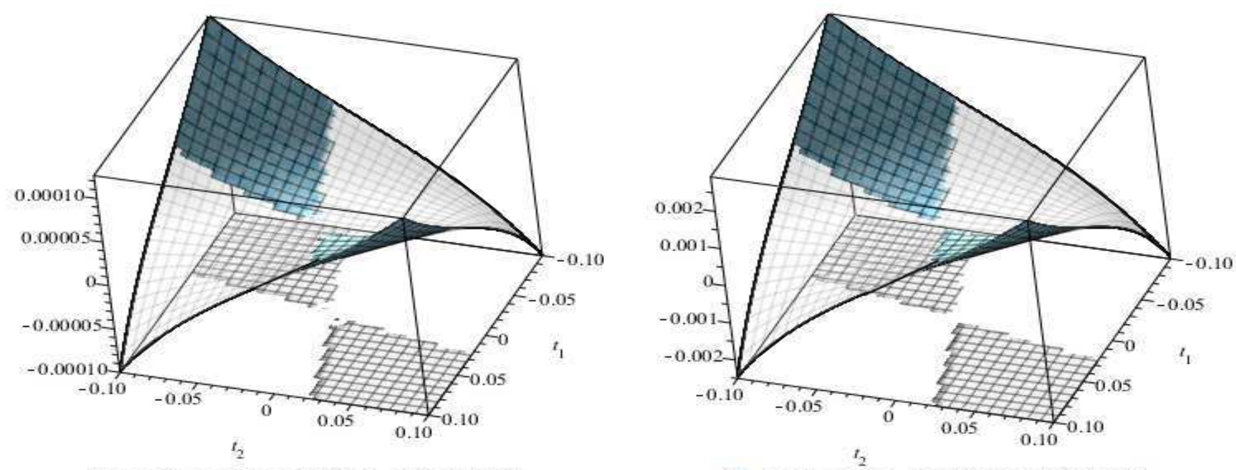


Figure 3.2: $\sigma_1 = 0.1$, $\sigma_2 = 0.105$, $\rho_1 = -0.5$, $\rho_2 = -0.2$, $\lambda = 0.5$



A: $\sigma_1 = 4$, $\sigma_2 = 4.1$, $\rho_1 = 0.333$, $\rho_2 = 0.2$, $\lambda = 0.99$

B: $\sigma_1 = 4$, $\sigma_2 = 4.1$, $\rho_1 = 0.333$, $\rho_2 = 0.2$, $\lambda = 0.5$

Figure 3.3

In Figure 3.4 we see an example where $\mathbb{S}$ is non-empty, and $\Lambda(\mathbf{t}) < 0$ for some $\mathbf{t} \in \mathbb{S}$. In other words, this is an example where the the EKFI holds true but does not yield an improvement relative to the standard KFI. In this example, matrix $\mathbf{C}_1 = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_1^2 \end{pmatrix}$ and $\mathbf{C}_2 = \begin{pmatrix} \sigma_2^2 & \sigma_2^2\rho_2 \\ \sigma_2^2\rho_2 & \sigma_2^2 \end{pmatrix}$. The value $\lambda = \lambda_1 = 0.99$.

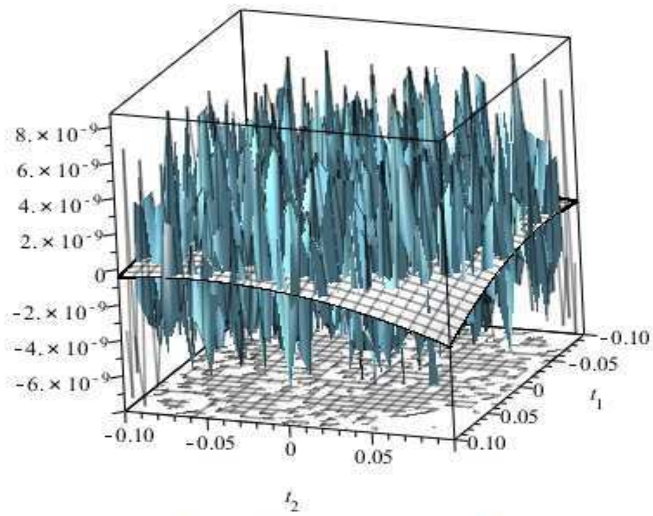
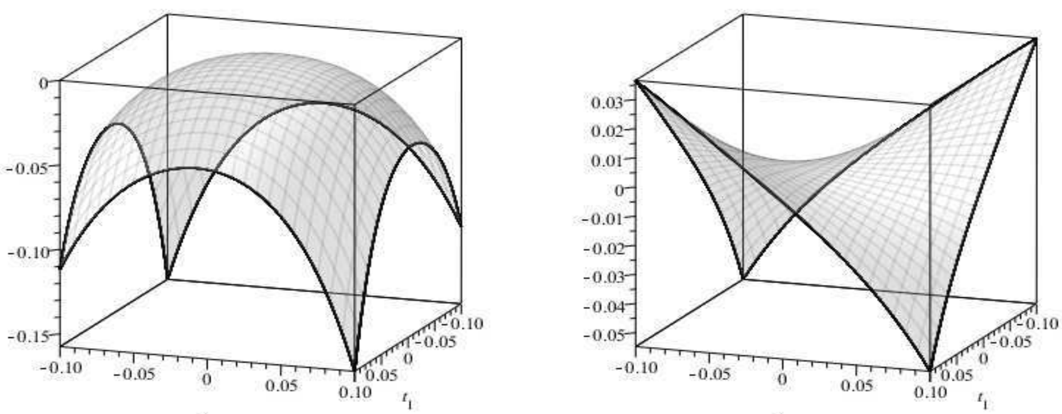


Figure 3.4: $\sigma_1 = 0.1$, $\sigma_2 = 0.102$, $\rho_2 = 0.01$, $\lambda = 0.99$



A: $\sigma_1 = 3$, $\sigma_2 = 5$, $\rho_1 = 0.333$, $\rho_2 = 0.2$, $\lambda = 0.5$ **B:** $\sigma_1 = 3.08$, $\sigma_2 = 4.05$, $\rho_1 = -0.666$, $\rho_2 = -0.2$, $\lambda = 0.5$

Figure 3.5

Finally, in Figures 3.5 the set $\mathbb{S}$ is empty. Consequently, the EKFI fails (within the indicated range of argument $\mathbf{t}$). Also, function Λ in Figure 3.5.A takes negative values for $\mathbf{t} \neq \mathbf{0}$ within the indicated range but Λ in Figure 3.5.B takes both negative and positive values within the indicated range. The matrices are $\mathbf{C}_1 = \begin{pmatrix} \sigma_1^2 & \sigma_1^2\rho_1 \\ \sigma_1^2\rho_1 & \sigma_1^2 \end{pmatrix}$ and $\mathbf{C}_2 = \begin{pmatrix} \sigma_2^2 & \sigma_2^2\rho_2 \\ \sigma_2^2\rho_2 & \sigma_1 2^2 \end{pmatrix}$ and $\lambda_1 = \lambda_2 = \lambda = 1/2$. The difference between these figures is that in Figure 3.5.A $\mathbf{t} = \mathbf{0}$ is a maximum of Λ (within the depicted range of $\mathbf{t}$) whereas in Figure 3.5.B the maximum is at the corner points.

Acknowledgements – YS thanks the Math Department, Penn State University, for the financial support and hospitality during the academic year 2014-5. SYS thanks the CAPES PNPD-UFSCAR Foundation for the financial support in the year 2014-5. SYS thanks the Federal University of Sao Carlos, Department of Statistics, for hospitality during the year 2014-5.

References

[1] M. Belis and S. Guiasu. A Quantitative and qualitative measure of information in cybernetic systems. *IEEE Trans. on Inf. Theory*, **14** (1968), 593–594.

[2] A. Clim. Weighted entropy with application. *Analele Universităţii Bucureşti, Matematică*, **Anul LVII** (2008), 223-231.

[3] T. Cover and J. Thomas. Determinant inequalities via information theory. *Siam J. Matrix Anal. Appl.*, **9** (1988), no. 3, 384–392.

[4] T. Cover and J. Thomas. *Elements of Information Theory.* New York: Wiley, 2006.

[5] A. Dembo, T. Cover and J.A. Thomas. Information-theoretic inequalities. *IEEE Trans. Inform. Theory*, **37** (6) (1991), 1501Ð1518.

[6] K. Fan. On a theorem of Weyl concerning eigenvalues of linear transformations, I. *Proc. Nat. Acad. USA,* **35** (1949), 652–655.

[7] K. Fan. On a theorem of Weyl concerning eigenvalues of linear transformations, II. *Proc. Nat. Acad. USA,* **36** (1950), 31–35.

[8] K. Fan. Maximum properies and inequalities for the eigenvalues of completely continuous operators. *Proc. Nat. Acad. USA,* **37** (1951), 760–766.

[9] S. Guiasu. Weighted entropy. *Report on Math. Physics*, **2** (1971), 165–179.

[10] M. Moslehian. Ky Fan inequalities. arXiv:1467v2, 2011.

[11] Y. Suhov and S. Yasaei Sekeh. Simple inequalities for weighted entropies. arXiv 1409.4102.